

1  
2  
3  
4  
5  
6  
7  
8  
  
9  
10  
11  
  
12  
13  
  
14  
15  
16

**BEFORE THE COMMONWEALTH OF KENTUCKY**  
**PUBLIC SERVICE COMMISSION**  
**SURREBUTTAL TESTIMONY OF ROBERT M. BELL, PH. D.**  
**ON BEHALF OF**  
**AT&T COMMUNICATIONS OF THE SOUTH CENTRAL STATES, INC.**  
**AND TCG OHIO, INC.**  
**CASE NO. 2001-105**  
**AUGUST 20, 2001**

**Q. PLEASE STATE YOUR NAME AND ADDRESS.**

A. My name is Robert M. Bell. My business address is AT&T Labs-Research, 180 Park Avenue, Florham Park, New Jersey 07932.

**Q. HAVE YOU PREVIOUSLY FILED TESTIMONY IN THIS DOCKET?**

A. Yes. I filed rebuttal testimony in this docket on July 9, 2001.

**Q. WHAT IS THE PURPOSE OF YOUR TESTIMONY?**

A. I respond to the rebuttal testimony of Dr. Edward J. Mulrow and to parts of the rebuttal testimony of Dr. William E. Taylor that deal with statistical issues.

1 **Q. IS DR. MULROW CORRECT WHEN HE STATES THAT YOUR MAIN**  
2 **ARGUMENT CONCERNING THE THIRD-PARTY TEST APPEARS TO**  
3 **BE THAT IT SHOULD HAVE USED THE SAME PRINCIPLES AS**  
4 **PERFORMANCE MEASUREMENT PLANS LIKE BELLSOUTH'S**  
5 **SEEM? (MULROW REBUTTAL P. 2).**

6 A. No. With the exception of how KCI treated benchmarks, each of my criticisms  
7 was derived from independent application of principles from statistical design and  
8 analysis. I criticize KCI's P-value analyses because they fail to consider both  
9 Type I and Type II errors. While those concepts are important for performance  
10 measurement also, I take into account the important distinction that sample sizes  
11 could be controlled in the third-party test while sample sizes cannot be controlled  
12 in the performance measurement setting. In particular, I recommend use of an  
13 "inconclusive" classification for third-party tests whenever the available evidence  
14 was insufficient to support either a "satisfied" or "not satisfied" conclusion.  
15 My other main criticisms— KCI's extensive use of professional judgment,  
16 implementation of military style testing, and the lack of blindness—all have no  
17 relationship to the performance measurement plans.

18 **Q. IS DR. TAYLOR CORRECT WHEN HE SUGGESTS THAT TRUNCATED**  
19 **Z HAS A BUILT-IN ASYMMETRY THAT FAVORS THE CLECS?**  
20 **(TAYLOR REBUTTAL PP. 43-44).**

21 A. No. Dr. Taylor's statement betrays a misunderstanding of truncated z. Negative  
22 values of truncated z provide evidence that BellSouth has discriminated.  
23 Consider a set of cells that produce a particular truncated z score. Suppose that  
24 we now find a new cell where BellSouth provided better service to the CLEC than  
25 it did to itself. If we add the new cell to the aggregation, what would happen to  
26 the truncated z score? If Dr. Taylor is correct—that BellSouth receives no credit

1 for cells of this type—then the truncated z score must either stay the same or go  
2 down (i.e., become more negative or less positive). But that is not what happens.  
3 Truncated z actually goes *up*, providing BellSouth with the credit that it is due.  
4 The formula for truncated z (page A-9 of EJM-1) is a ratio with three pieces:

5 
$$Z^T = \frac{\sum_j W_j Z_j^* - \sum_j W_j E(Z_j^* | H_0)}{\sqrt{\sum_j W_j^2 \text{Var}(Z_j^* | H_0)}}$$

6 Each of those pieces is a sum over the cells being aggregated, with the subscript *j*  
7 indexing the cells. Consider what happens when we add to the aggregation the  
8 new cell described above, where BellSouth provided better service to the CLEC  
9 than it did to itself. Because BellSouth provided better-than-expected service to  
10 the CLEC,  $Z_j^* = 0$  for the new cell; consequently the first sum in the numerator  
11 does not change. However, the other two sums do change when any new cell is  
12 included in the aggregation. Because  $E(Z_j^* | H_0)$  is always negative, the second  
13 sum in the numerator must decrease and the entire numerator must *increase* when  
14 we include the new cell. Similarly, because  $\text{Var}(Z_j^* | H_0)$  is always positive, the  
15 denominator must also *increase*. In the case of interest, when truncated z is  
16 negative ( $Z^T < 0$ ), adding the new cell makes the numerator less negative and the  
17 denominator more positive, leading to an increase in the value of truncated z.  
18 That is, BellSouth does receive a credit when it provides better-than-expected  
19 service to a CLEC.

1 **Q. DR. MULROW REFERS TO THE USE OF OTHER FORMS OF**  
2 **STATISTICAL AGGREGATION IN SOME STATES TO SUPPORT**  
3 **BELLSOUTH'S PROPOSAL TO USE TRUNCATED Z IN KENTUCKY.**  
4 **(MULROW REBUTTAL P. 6). ARE THOSE OTHER USES RELEVANT**  
5 **TO THE KENTUCKY PLAN?**

6 A. I do not believe that they are. Both BellSouth and the CLECs have accepted the  
7 principle of balancing Type I and Type II errors in Kentucky. I do not believe  
8 that balancing is used in any of the states that Dr. Mulrow mentions. In particular,  
9 the "K-value" method completely ignores Type II errors. My concern about  
10 truncated z is that it should not be used to aggregate heterogeneous cells because  
11 it interferes with proper balancing.

12 **Q. IS DR. TAYLOR CORRECT IN ASSERTING THAT THE**  
13 **CONCLUSIONS YOU DRAW ABOUT MATERIALITY FROM YOUR**  
14 **TABLE 1 ARE ARTIFACTS OF THE STANDARD DEVIATION THAT**  
15 **YOU ASSUME FOR ORDER COMPLETION INTERVAL? (TAYLOR**  
16 **REBUTTAL PP. 59-60).**

17 A. Not exactly. He is correct that the conclusions are sensitive to assumptions about  
18 the standard deviation of the variable. The point of my table is to illustrate how  
19 large a disparity is deemed to be material by alternative values of delta. Because  
20 those disparities are proportional to BellSouth's standard deviation, the assumed  
21 value directly affects the conclusions from my example.  
22 However, Dr. Taylor then proceeds to discuss what the table would look like with  
23 a standard deviation of 0.5 days—an absurdly small value. I assumed a standard  
24 deviation of 5 days because distributions for waiting times tend to have long tails  
25 (i.e., some customers may take 30 days or more to provision). Consequently,  
26 these measures would be expected to have standard deviations at least as large as  
27 their means. Data from the BellSouth Monthly State Summary report for

1 Kentucky in June 2001 support my position. The report lists BellSouth means  
2 and standard deviations of Order Completion Interval (measure P-4, page 2) for  
3 18 cells with two or more BellSouth observations. For 13 of 18 cells, the  
4 BellSouth standard deviation exceeds the mean. For example, for 4,859  
5 dispatched residential customers with fewer than ten circuits, the BellSouth mean  
6 interval was 6.58 days, while the BellSouth standard deviation was 8.654—32  
7 percent higher than the mean. In half the cells, the ratio of the standard deviation  
8 to the mean is 1.3 or higher. Consequently, Table 1 probably *understates* the  
9 sizes of disparities implied by the tabled values of delta. In contrast, the standard  
10 deviation proposed Dr. Taylor, based on a ratio of 0.1, is totally indefensible.

11 **Q. HOW DO YOU RESPOND TO DR. TAYLOR'S CHARGE THAT YOUR**  
12 **CHARACTERIZATION OF DIFFERENCES IN TABLE 2 IS**  
13 **MISLEADING? (TAYLOR REBUTTAL P. 61).**

14 A. I disagree. Rather than comparing 1% bad service for BellSouth customers with  
15 5% bad service for CLEC customers and concluding that the CLEC rate is five  
16 times the BellSouth rate, Dr. Taylor advocates the alternative presentation that the  
17 CLEC satisfaction rate is 96% of BellSouth's satisfaction rate. When describing  
18 rare events, however, I believe that comparing the probabilities of non-events is  
19 the deceptive practice. I can imagine a cigarette ad aimed at current smokers:  
20 "Don't bother trying to quit. If you keep smoking, you will still have 96 percent  
21 as good a chance of not getting lung cancer as if you had quit." Dr. Taylor does  
22 make a valid point that whether an increase from 1 percent to 5 percent is material  
23 depends on the seriousness of the event. The Commission should consider  
24 whether a five-fold increase would be material for the types of bad service event

1 that occur about 1 percent of the time for BellSouth customers. Examples of this  
2 in the June data for Kentucky are: Customer Trouble Reports (M&R-2)  
3 residential/non-dispatched; Customer Trouble Reports (M&R-2)  
4 ISDN/dispatched; Jeopardies-Mechanized (P-2) two-wire analog loop/non-design;  
5 and Average Response Interval (OSS-4) > 10 seconds/CRIS/Region. If five-fold  
6 increases in the frequency of these problems would be material, then delta should  
7 be set no higher than 0.25.

8 **Q. WHAT IS YOUR RESPONSE TO DR. TAYLOR'S CONCERN THAT**  
9 **TABLES 1 AND 2 DO NOT ADDRESS THE ECONOMIC OR MATERIAL**  
10 **SIGNIFICANCE OF DISPARITIES, AND, CONSEQUENTLY, THEY DO**  
11 **NOT HELP TO FIND A DELTA THAT BALANCES BELL SOUTH'S**  
12 **COMMERCIAL GAIN FROM DISCRIMINATION WITH ITS RISK OF**  
13 **PAYING A PENALTY WHEN IT DOES NOT DISCRIMINATE?**  
14 **(TAYLOR REBUTTAL P. 62).**

15 A. None of the developers of the balancing critical value methodology has ever  
16 suggested that one of its objectives was to balance BellSouth's commercial gain  
17 from discriminating against any other type of risk. Accordingly, that is not an  
18 objective when setting delta.

19 Concerning materiality, I have never claimed that the tables alone imply what  
20 delta should be. I have offered the tables as tools to aid the Commission in  
21 interpreting the implications of alternative values of delta.

22 **Q. WITH RESPECT TO YOUR TABLE 2, DR. MULROW COMMENTS,**  
23 **"BELLSOUTH DOES NOT USE 'DELTA' TO DEFINE THE**  
24 **ALTERNATIVE HYPOTHESIS FOR PROPORTION MEASURES."**  
25 **(MULROW REBUTTAL PP. 9-10). DOES THAT MEAN THAT YOUR**  
26 **TABLE IS IRRELEVANT?**

27 A. No. First, both BellSouth and the CLECs do use delta for mean measures, so the  
28 Commission needs to select a value of delta for at least those measures. Although

1 Table 1 is more directly relevant, I believe that Table 2 also is a useful tool for  
2 evaluating alternative values of delta. Second, the CLEC plan uses delta for  
3 proportion measures, so Table 2 becomes directly relevant if the Commission uses  
4 the CLECs' proposed method for proportions.

5 **Q. AS DR. MULROW NOTES, BELLSOUTH PROPOSES USING THE**  
6 **"ODDS RATIO" FOR BALANCING WITH PROPORTION MEASURES.**  
7 **(MULROW REBUTTAL PP. 10-11). IF THE COMMISSION USES**  
8 **BELLSOUTH'S METHOD WOULD IT NEED TO SPECIFY A VALUE OF**  
9 **THE ODDS RATIO FOR ALTERNATIVE HYPOTHESES?**

10 A. Yes, for Dr. Mulrow's proposed method to be complete, in addition to delta, the  
11 Commission would need to specify this second parameter. However, to the best  
12 of my knowledge, BellSouth has not proposed a value for this parameter.

13 **Q. YOU AND DR. MULROW DISAGREE ON HOW "MATERIALITY"**  
14 **SHOULD BE DEFINED. (MULROW PP. 17, 19-21). PLEASE EXPLAIN**  
15 **THAT DIFFERENCE.**

16 A. Dr. Mulrow states that materiality corresponds to a disparity of one-half delta  
17 times BellSouth's standard deviation. I base my definition on the principle  
18 behind balancing, that the probability of a Type I error assuming parity should  
19 equal the probability of a Type II error assuming a material disparity.  
20 Consequently, materiality refers to the size of the disparity specified in the  
21 alternative hypothesis—delta times the BellSouth standard deviation. Dr. Taylor  
22 seems to concur with my opinion. In Footnote 19 on p. 42, he states, "Materiality  
23 must be used to determine the degree of discrimination or performance disparity  
24 at which it is appropriate to balance Type I and II error probabilities."

1 **Q. IN SUPPORT OF YOUR DEFINITION, YOU QUOTED THE LOUISIANA**  
2 **STATISTICIAN’S REPORT. HOWEVER, DR. MULROW RESPONDED**  
3 **THAT HE DOES NOT BELIEVE IT WAS THE INTENTION OF THE**  
4 **AUTHORS, INCLUDING HIMSELF, “TO MAKE MATERIALITY**  
5 **SYNONYMOUS WITH THE VALUE OF ‘DELTA’.” (MULROW**  
6 **REBUTTAL P. 20). DOES DR. MULROW’S RESPONSE MAKE SENSE**  
7 **TO YOU?**

8 A. No. If the authors had intended for remedies to begin when the observed disparity  
9 (weighted, if necessary) became material, they could have done that much more  
10 simply, without getting into Type I and Type II errors. Indeed, if that had been  
11 their intention, what is it that they were “balancing”? To me, the only logical  
12 explanation is that the authors were balancing Type I error under parity with Type  
13 II error for a material disparity.

14 **Q. PLEASE RESPOND TO DR. TAYLOR’S CLAIM THAT YOU IGNORE**  
15 **THE SALIENT CHARACTERISTIC OF TESTING WITH BALANCING.**  
16 **(TAYLOR P. 56).**

17 A. Dr. Taylor’s statement severely distorts the impact of delta. In lines 13-14, Dr.  
18 Taylor writes, “So, if a large delta, particularly with large samples, seems to lower  
19 the Type I error rate almost to zero (which favors BellSouth) . . .” This part is  
20 correct; as delta increases, Type I error decreases. However, he continues the  
21 sentence: “. . .then so does it lower the Type II error rate almost to zero (which  
22 favors CLECs).” The last part of the sentence is true only because Dr. Taylor is  
23 referring to an alternative hypothesis that also changes with delta. This is like  
24 saying that my chance of hitting a target increases as the target gets farther  
25 away—as long as I keep using a larger target. The appropriate discussion would  
26 use a fixed alternative hypothesis. For any fixed alternative hypothesis, the Type  
27 II error rate *increases* as delta increases. The absurdity of Dr. Taylor’s analysis is

1 made evident by his parenthetical phrases, which say that a large delta favors  
2 BellSouth *and* favors CLECs. If that were the case, both sides would be asking  
3 for a delta value of 20.

4 **Q. IS DR. TAYLOR'S CRITICISM OF THE CLEC PLAN FOR USING THE**  
5 **MODIFIED Z STATISTIC IN DETERMINING REMEDY PAYMENTS**  
6 **WARRANTED? (TAYLOR REBUTTAL P. 45).**

7 A. No. Dr. Taylor goes overboard in his criticism. His claim that a statistical  
8 decision rule may not be used for determining the severity of material  
9 performance violations is like saying that one cannot use a steak knife to cut an  
10 apple because it was not designed for that purpose. For a fixed sample size, the z-  
11 score for a mean is proportional to the size of the observed disparity in the means.  
12 Clearly, it is a measure of the severity of the violation. Dr. Taylor's criticism is  
13 also puzzling because BellSouth's remedy calculation uses a z score, truncated z,  
14 as part of its calculation.

15 **Q. IN YOUR REBUTTAL TESTIMONY, YOU CRITICIZED BELLSOUTH'S**  
16 **"AFFECTED VOLUME" CALCULATION. IN HIS REBUTTAL, DR.**  
17 **MULROW EXPLAINS THE CONCEPT BEHIND THE CALCULATION.**  
18 **(MULROW REBUTTAL PP. 21-22). DID HE ALLAY YOUR CONCERN?**

19 A. No, he confirmed it. My criticism was that instead of indicating how far  
20 BellSouth is from parity, BellSouth's calculation approximates how far BellSouth  
21 is from not getting caught. Dr. Mulrow confirmed my impression: "under  
22 BellSouth's plan, a calculation is made of the number of transactions that would  
23 have had to be accomplished more quickly (if the time interval was the relevant  
24 measure) *in order to avoid having a failure.*" (Mulrow Rebuttal p. 21 (emphasis

1 added)). This calculation is inappropriate, because the goal of the  
2 Telecommunications Act is parity service.

3 **Q. DR. MULROW ALSO REFERS TO AN ANALYSIS SHOWING THAT**  
4 **THE AFFECTED VOLUME CALCULATION IS ALWAYS AT LEAST AS**  
5 **LARGE AS THAT CALCULATED BY A LINEAR PROGRAM.**  
6 **(MULROW REBUTTAL PP. 25-27). DID THAT FINDING ALLAY YOUR**  
7 **CONCERN?**

8 A. No. The linear program solution is still based on the flawed concept of just barely  
9 getting out of violation. The exact answer to the wrong question tells us nothing.

10 **Q. DOES THIS CONCLUDE YOUR TESTIMONY?**

11 A. Yes.